Spectral estimation of terahertz time-domain signals using a simple parametric method

F. Sanjuan✉ and C. Galarza

A simple method for spectral estimation of terahertz signals using an autoregressive model is developed. The technique shows a better spectral variance and resolution over the use of the Fast Fourier Transform, which is commonly utilised in this kind of signal. The results are validated comparing the water vapour absorption lines and the standard deviation of a set of 24 measurements between methods. As a consequence of improving the signal variance, a better signal-to-noise ratio is achieved without enhancing the experiment. Such result also produces an increase in the signal bandwidth and less error when optical parameters are estimated.

Introduction: Terahertz time-domain spectroscopy (THz-TDS) is a technique used to obtain the optical parameters of a given material in the THz band. Mainly it is of interest to the medical, pharmaceutical, security, and industrial sectors because it performs non-invasive tests on the materials under study, it is non-ionising, and also because the THz band holds material spectral signatures and physical processes.

THz-TDS usually produces temporal pulses of about one pico-second duration that generate a broadband response in the frequency domain. The material optical parameters are typically obtained using the information of the absolute value and phase of the transfer function formed by the ratio of the spectral THz signal measured with and without a sample of the material under test.

In general, the spectral information is obtained by applying the fast Fourier transform (FFT), which is a poor non-parametric spectral estimator. The main limitations of this technique are summarised below [1, 2]:

(i) The frequency resolution depends on the inverse of the temporal length of the measured signal. Increasing the measured length is not practical or even possible experimentally because of the time required and the delay line limitations. It should be noted that zero padding in the temporal signal does not generate new spectral information.
(ii) The FFT assumes signal periodicity causing discontinuities in the 'periods' and generating spectral leakage. To minimise this effect, tapered windows are implemented at the expense of smoothing the signal and losing the frequency resolution.
(iii) The variance of the estimated power spectral density (PSD) does not improve when the number of samples is increased.

To solve the limitations mentioned above, other techniques such as parametric methods for rational spectra are used [2]. Models such as autoregressive moving average (ARMA), autoregressive (AR), or moving average are utilised for representing the signal spectra. In [3], an ARMA model was used to estimate the material properties using THz time-domain data. Nevertheless the method loses simplicity because identifying ARMA models require solving a non-convex optimisation problem, even when the model order is set.

In this Letter, a different parametric spectral estimation method is proposed. In particular, we obtained the THz PSD using an AR model. Unlike ARMA models, its identification is obtained by solving a convex optimisation problem that guarantees an optimum result for a given order. It is known that AR spectral estimation improves on FFT limitations. It assumes no signal periodicity, and the spectral resolution does not depend on the signal length [1]. Moreover, it is a consistent estimator and the variance of the spectral estimate improves when the number of samples increases [4].

In this Letter, we show those improvements by processing experimental data from two different THz-TDS experiments. First, we analyse the estimator spectral variance by comparing the standard deviation of a set of 24 measurements between the two methods. Then, we compare the water vapour absorption lines to exhibit the improvement in the frequency resolution. In the first experiment, we show that the parametric estimate leads to a larger signal-to-noise ratio (SNR) than the traditional FFT based due to the decrease in the estimated variance. On the other hand, in the second experiment, we obtain a more accurate estimate from the parametric technique because of the improvement in the frequency resolution.

In the sequel, we present the estimation method that has been used and then we show experimental results.

Method: To obtain the signal spectra, we model the temporal measurements $[y(t)]$ from the THz-TDS as a result of the convolution of white noise $[e(t)]$ with an AR system as shown in Fig. 1 [2], where $A(\omega) = 1 + a_1 e^{-i\omega} + a_2 e^{-i2\omega} + \cdots + a_n e^{-in\omega}$, n is the model order, a_k are the model coefficients, and σ^2 is the white noise variance. The PSD of $y(t)$ results in

$$\phi_y(\omega) = \left|\frac{1}{A(\omega)}\right|^2 \sigma^2, \quad \omega \in [0,\ 2\pi] \tag{1}$$

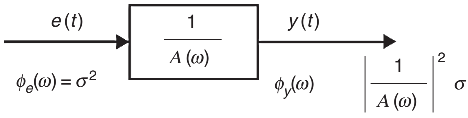

Fig. 1 *Signal modelling*

To estimate $\phi_y(\omega)$, n, a_k, and σ^2 must be identified from the measured signal $y(t)$.

Given a model order n, the whole issue is summed up to estimate the model parameters.

From Fig. 1, we obtain

$$y(t) = -\sum_{k=1}^{n} a_k y(t-k) + e(t) \tag{2}$$

Multiplying (2) by $y(t–i)$ and taking its expectation, the Yule Walker equations are obtained

$$r_y(i) + \sum_{k=1}^{n} a_k r_y(i-k) = 0 \quad \text{for} \quad i > 0 \tag{3}$$

$$r_y(0) + \sum_{k=1}^{n} a_k r_y(-k) = \sigma^2 \quad \text{for} \quad i = 0 \tag{4}$$

where $r_y(i) = E\{y(t)\,y(t-i)\}$ is the autocorrelation of the process $y(t)$. When the signal $y(t)$ is measured over a finite interval of time $T = NT_s$, where T_s is the sampling period, and N is the number of samples, the time lag i in (3) takes values in $[1, N]$. Then, (3) and (4) produce an $(N+1) \times (n+1)$ linear system equation that can be written as

$$Ru = v \tag{5}$$

where $u = [1,\ a_1,\ \ldots,\ a_n]^T \in \mathbb{R}^{n+1}$, $v = [-r(0) + \sigma^2, -r(1), \ldots, -r(N)]^T \in \mathbb{R}^{N+1}$, and $R \in \mathbb{R}^{(N+1)\times(n+1)}$ the covariance matrix. When $N > n$, the pseudo-inverse of $\boldsymbol{R}$ gives the least-square solution of (5), which provides the model coefficients and the white noise variance used to estimate $\phi_y(\omega)$ (1) [2].

To ensure the selection of an appropriate order, we first solve the Yule Walker equations for different values of n ranging between 5 and 55% of the number of measured samples [1]. For each model order we compute the error

$$e(t) = y_m(t) - \hat{y}(t) \tag{6}$$

where $y_m(t)$ is the measured signal and $\hat{y}(t)$ is the estimated temporal signal obtained using (2). We keep the smallest order whose estimation error (6) is approximately uncorrelated in time.

To quantify the goodness of fit, we compute the coefficient of determination ($\boldsymbol{R}$-squared) between $y_m(t)$ and $\hat{y}(t)$, which is defined as

$$\boldsymbol{R}^2 = 1 - \frac{\left(y_m(t) - \hat{y}(t)\right)^2}{\left(y_m(t) - \bar{y}\right)^2} \tag{7}$$

where $\bar{y}$ is the mean of the measured data. An $\boldsymbol{R}$-squared value of 1 indicates a perfect fit. Experimentally, we will consider that $\boldsymbol{R}^2 = 0.99$ is good enough. To avoid pernicious overfitting, the minimum number of coefficients that fulfils both conditions must be used. In practice we have observed that when one condition is accomplished, also it is the other one.

Implementation: To analyse the improvement in the spectral variance, we performed a set of 24 measurements with the least amount of time between them, to compare the standard deviation between the parametric method and the FFT based.

We have used a transmission THz-TDS from the Nanophotonics Technology Center (Valencia, Spain) formed by a mode-locked laser (FemToFerb, Toptica) and photoconductive antennas (TERA15, Menlo Systems).

The data acquisition was made by a lock-in amplifier (SR830, Stanford Research Systems) with a time delay constant set in 200 ms. The sample time was 100 fs and the temporal length 250 ps. The sample used was a silicon wafer of 640 µm width.

Using our model selection scheme for a particular realisation of the experiment, we have obtained $n = 125$. In Fig. 2*a*, we show the autocorrelation of $e(t)$ (6), which is approximately a $\delta(t)$, and in Fig. 2*b* we show the temporal and estimated signal. An $\boldsymbol{R}$-squared of 0.9988 was computed, showing that both conditions are well accomplished. After following the same procedure for the 24 realisations, we have obtained the same model order.

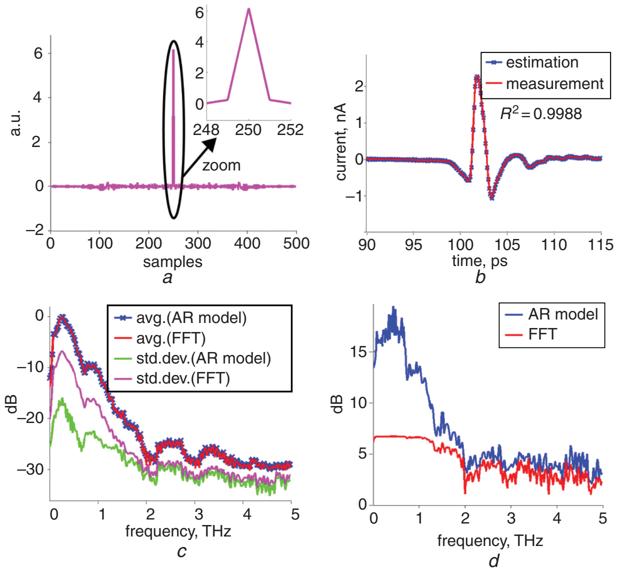

Fig. 2 *Temporal and spectral results*

a $e(t)$ autocorrelation
b Measured and estimated temporal signals
c Spectral averages and standard deviations
d Spectral SNRs

Finally, by estimating each spectra making the square root of (1) with all the AR models obtained, and also by applying the absolute value of the FFT (rectangular window), we have computed the averages and standard deviations.

In Fig. 2*c* the results are shown, where the spectral averages are very similar but the standard deviation of the proposed method shows an improvement of about 10 dB over the FFT. This results in a better SNR and also in an increased signal bandwidth (see Fig. 2*d*) due to the possibility of detecting smaller variations at high frequencies. The SNR was computed by making the ratio between the spectral average over the standard deviation.

To analyse the improvement in the spectral resolution, we compare the water vapour absorption lines published in the literature [5, 6], with those obtained using both methods. The measurements were made with a free space transmission THz-TDS from the Research Optics Center (La Plata, Argentina) formed by a mode-locked laser (MAI-TAI, Spectra Physics) and photoconductive antennas (Teravils). The data was obtained with a lock-in amplifier (SR830, Stanford Research Systems), with a time delay constant of 30 ms. The sample time was 280 fs and the temporal length 230 ps.

Applying the proposed method, we obtained a model order of $n = 821$. The FFT results were obtained by zero padding in the temporal signal to approach as much as possible the absorption lines to those presented in the literature.

Table 1 shows the values obtained through both procedures and those published in the literature, showing a better result when an AR model is used.

Table 1: Comparison of THz water vapour absorption lines (THz)

Published lines [5, 6]	AR model	FFT
0.557	0.5579	0.5618
0.988	0.9872	0.9870
1.0977±0.0005	1.0985	1.0987
1.16440±0.0001	1.1653	1.1657

Conclusion: In this Letter, we have proposed a simple parametric spectral estimation method for obtaining the PSD of THz time-domain signals achieving a better result than the FFT-based estimation. A better spectral variance was obtained, leading to a better SNR without improving the experiment. Such result gives the chance of detecting a wider bandwidth due to the possibility of detecting smaller changes at higher frequencies. Moreover, this allows obtaining the material optical parameters with a minor error when the Kramer–Kronig relations are used.

In addition, we have also shown improvements in the frequency resolution, which give the possibility of obtaining a more accurate spectral result.

Acknowledgments: The authors thank the contributions and collaborations of Borja Vidal, Alexander Bockelt, Pablo Marchi, Raymundo Albert, and Francisco Messina. This work was partially supported by a PME-2006-00018 (Equipment Modernization Project for research laboratories, 2006) grant from Agencia Nacional de Investigaciones Científicas y Tecnológicas del Ministerio de la Investigación, Ciencia e Innovación de la Nación (Argentina).

Submitted: *20 August 2015*
doi: 10.1049/el.2015.2955
One or more of the Figures in this Letter are available in colour online.

F. Sanjuan (*CSC-CONICET, Godoy Cruz 2390, AR-C1425FQD Buenos Aires, Argentina*)

✉ E-mail: federico.sanjuan@conicet.gov.ar

C. Galarza (*CSC-CONICET, Godoy Cruz 2390, AR-C1425FQD Buenos Aires, Argentina*)

C. Galarza: Also with Universidad de Buenos Aires, Avenue Paseo Colón 850, AR-C1063ACV Buenos Aires, Argentina

References

1 Schlindwein, F.S., and Evans, D.H.: ‘Autoregressive spectral analysis as an alternative to fast Fourier transform analysis of Doppler ultrasound signals’, in Arnold, E. (ED.): ‘Diagnostic vascular ultrasound’ (Hodder Arnold Publication, CRC Press, 1992), vol. 8, pp. 74–84
2 Stoica, P., and Moses, R.: ‘Spectral analysis of signals’ (Prentice-Hall, USA, 2005)
3 Tych, K.M., Wood, C.D., and Tych, W.: ‘A simple transfer function based approach for estimating material parameters from Terahertz time-domain data’, *IEEE Photonics J.*, 2014, **6**, pp. 1–11
4 Ljung, L.: ‘System identification’ (Prentice-Hall, USA, 1999)
5 Van Exter, M., Fattinger, Ch., and Grischkowsky, D.: ‘Terahertz time-domain spectroscopy of water vapor’, *Opt. Lett.*, 1989, **14**, pp. 1128–1130
6 Cheville, R.A., and Grischkowsky, D.: ‘Far-infrared foreign and self-broadened rotational linewidths of high-temperature water vapor’, *J. Opt. Soc. Am. B*, 1999, **16**, pp. 317–322